Robust Low-Rank Latent Feature Analysis for Spatiotemporal Signal Recovery

Di Wu, *Member, IEEE*, Zechao Li, *Senior Member, IEEE*, Zhikai Yu, *Student Member, IEEE*, Yi He, *Member, IEEE*, and Xin Luo, *Senior Member, IEEE*

***Abstract*— Wireless sensor network (WSN) is an emerging and promising developing area in the intelligent sensing field. Due to various factors like sudden sensors breakdown or saving energy by deliberately shutting down partial nodes, there are always massive missing entries in the collected sensing data from WSNs. Low-rank matrix approximation (LRMA) is a typical and effective approach for pattern analysis and missing data recovery in WSNs. However, existing LRMA-based approaches ignore the adverse effects of outliers inevitably mixed with collected data, which may dramatically degrade their recovery accuracy. To address this issue, this article innovatively proposes a latent feature analysis (LFA) based spatiotemporal signal recovery (STSR) model, named LFA-STSR. Its main idea is twofold: 1) incorporating the spatiotemporal correlation into an LFA model as the regularization constraint to improve its recovery accuracy and 2) aggregating the L_1-norm into the loss part of an LFA model to improve its robustness to outliers. As such, LFA-STSR can accurately recover missing data based on partially observed data mixed with outliers in WSNs. To evaluate the proposed LFA-STSR model, extensive experiments have been conducted on four real-world WSNs datasets. The results demonstrate that LFA-STSR significantly outperforms the related six state-of-the-art models in terms of both recovery accuracy and robustness to outliers.**

***Index Terms*— Data recovery, L_1-norm, latent feature analysis (LFA), low-rank matrix approximation (LRMA), robust representation learning, spatiotemporal correlation, wireless sensor networks (WSNs).**

NOMENCLATURE

M, N	Number of sensors and time slots, respectively.
Y	Original incomplete data matrix from WSNs, $Y \in \mathbb{R}^{M \times N}$.
Y_K, Y_U	Known and unknown entry sets of Y, respectively.
W	Weight matrix comes from spatial information of sensors.
L	Laplacian matrix comes from spatial information of sensors.
$y_{i,j}$, $w_{i,j}$	W's and Y's element at ith row and jth column, respectively.
$\hat{y}_{i,j}$	Estimation of $y_{i,j}$.
D	Temporal differential matrix.
S, U	Latent feature matrices with size $M \times k$ and $N \times k$, respectively.
$s_{i,.}$, $u_{j,.}$	ith row-vector and jth row-vector of S and U, respectively.
k	Latent feature dimension of S and U.
J	$M \times N$ binary index matrix.
$\varepsilon(S, U)$	Objective function with respect t S and U.
$\varepsilon_{i,j}$	Instant state of $\varepsilon(S, U)$ on a single entry $y_{\mathrm{i,j}}$.
z_1, z_2	Hyperparameters of controlling the effects of spatial and temporal smoothness, respectively.
T	Maximum iteration number.
t	tth iteration, $t \in \{1, 2, \ldots, T\}$.
β_1, β_2	Aggregation weights of controlling the effects of L_1 and L_2 norms, respectively.
$\beta t1$, $\beta t2$	State of β_1 and β_2 at the tth iteration, respectively.
O_1^t, O_2^t, $O_{1,2}^t$	Partial estimation errors depend on L_1-norm, L_2-norm, and L_1-L_2-norm at the tth iteration, respectively.
B_{O1}^t, B_{O2}^t, $B_{O1,2}^t$	Cumulative loss corresponding to O_1^t, O_2^t, $O_{1,2}^t$ till the tth iteration, respectively.
λ	Regularization hyperparameter.
η	Learning rate.
σ	Balance coefficient controlling ensemble of L_1-norm and L_2-norm.

Manuscript received 31 October 2022; revised 11 July 2023; accepted 1 December 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62176070 and Grant 62272078, in part by the Science and Technology Foundation of State Grid Corporation of China under Grant 1400-202357341A-1-1-ZN, and in part by the Chongqing Technical Innovation and Application Development Special Project under Grant CSTB2023TIAD-KPX0037. *(Corresponding author: Xin Luo.)*

Di Wu and Xin Luo are with the College of Computer and Information Science, Southwest University, Chongqing 400715, China (e-mail: wudi.cigit@gmail.com; luoxin@swu.edu.cn).

Zechao Li is with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China (e-mail: zechao.li@njust.edu.cn).

Zhikai Yu is with the School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: yuzhikai@cqupt.edu.cn).

Yi He is with the Department of Computer Science, Old Dominion University, Norfolk, VA 23529 USA (e-mail: yihe@cs.odu.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2023.3339786.

Digital Object Identifier 10.1109/TNNLS.2023.3339786

I. INTRODUCTION

WIRELESS sensor networks (WSNs) are task-oriented kind of networks. A WSN consists of multiple sensor nodes to monitor the desired information in real time.

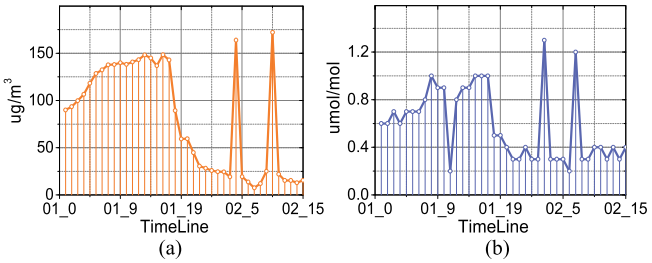

Fig. 1. Data distribution of real Beijing air quality. (a) PM2.5. (b) CO.

The monitored information is processed by embedded computing modules first and then transmitted to remote users or stations through a wireless communication network. With the rapid development of the Big Data Era, WSN has been widely applied in various fields, like smart city [1], [4], [31], industrial production [2], energy [3], cyber security [44], bridge engineering [5], transportation [6], [43], and computer vision [30], etc. It is becoming an emerging and promising developing area in the intelligent sensing field.

Commonly, the collected data from WSNs are incomplete due to various reasons, e.g., sudden sensors breakdown, saving energy by deliberately shutting down partial nodes, human errors, uncontrollable factors, etc. [7]. However, WSNs are usually deployed to the designed locations to collect data with the preset time interval for a long time, making spatiotemporal correlation inhere in the collected incomplete data from WSNs [8], [9], [10], [11], [12]. For example, two temperature sensors with close locations tend to have similar temperature data in a small time range (e.g., 1 min). On this basis, the incomplete WSNs data can be modeled into an incomplete matrix, where the rows represent sensor nodes, the columns represent time slots, and the entries represent the recorded values of WSNs data. Such a matrix is low rank because its rows and columns have high spatiotemporal correlation. Hence, low-rank matrix approximation (LRMA) is widely employed to recover the missing data of WSNs [32].

Matrix factorization is one of the most popular LRMA models due to its efficiency and effectiveness [13], [14], [15], [16], [70]. It trains two latent factor matrices to estimate the missing data of a low-rank incomplete matrix from WSNs based on the observed data only. However, existing related approaches ignore the adverse effects of outliers that may dramatically degrade their recovery accuracy. Unfortunately, the outliers are inevitably mixed with the collected data from WSNs in real scenarios. To be more intuitive, Fig. 1 shows a real example to illustrate this issue.

Example 1: We select two real-world datasets of average PM2.5 and CO concentration of one hour in Beijing, China, from the Beijing Municipal Ecological and Environmental Monitoring Center at link https://quotsoft.net/air/. Fig. 1 shows the data distribution during the period from 0 o'clock May 1st, 2014 to 14 o'clock May 2nd, 2014. Although PM2.5/CO tends to have different values at different times, most values fall into a normal range. For example, the PM2.5 values mainly fall in the interval of [3.0, 150.0]. However, there are also some values deviating far from the normal range. For instance, Fig. 1(a) shows that the PM2.5 values at times of 4 o'clock and 10 o'clock on May 2nd, 2014, are 164 and 172.2, which are far beyond the normal range. In general, such kind of PM2.5 values should be treated as outliers.

Example 1 substantiates the unavoidability of outliers in collected data from WSNs. To eliminate the adverse effects of outliers, a straightforward way is to delete them directly. However, such a way is not always reasonable because the outliers are not always equal to the noises in real applications. For example, a PM2.5 outlier may be caused by factories' sudden waste gas emissions. Needless to say, such an outlier cannot be deleted directly. Alternatively, can we build a robust LRMA model that can minimize the adverse effects of outliers during data recovery in WSNs?

In response, this article proposes a latent feature analysis (LFA) based spatiotemporal signal recovery (STSR) model, named LFA-STSR, to recover missing data based on partially observed data mixed with outliers in WSNs. Its main idea is twofold: 1) incorporating the spatiotemporal correlation into an LFA model as the regularization constraint to improve its recovery accuracy and 2) aggregating the L_1-norm into the *Loss* part of the LFA model to improve its robustness to outliers. With such a design, it possesses the merits of both high accuracy and robustness during data recovery. The main contributions of this work are as follows.

1) An LFA-STSR model is proposed. It can accurately recover missing data based on partially observed data mixed with outliers in WSNs.
2) Theoretical analyses and proofs are provided for the proposed LFA-STSR model.
3) The implementation algorithm is designed for the proposed LFA-STSR model.

Besides, extensive experiments on four real-world WSNs datasets are performed to evaluate the proposed LFA-STSR model, including comparison with the state-of-the-art models and analyses of its characteristics. The results verify that LFA-STSR has significantly higher recovery accuracy than the comparison models, especially when the incomplete data have outliers.

II. Related Work

The LRMA model treats the data recovery of WSNs as a low-rank matrix recovery problem, where the rows represent sensor nodes, the columns represent time slots, and the entries represent the recorded values of WSNs data. LRMA estimates the missing data of an incomplete low-rank matrix based on its observed data. Recently, many sophisticated LRMA-based models have been proposed to achieve good performance in data recovery, including graph regularized [17], nonnegative constrained [18], generalized separable and nonnegative constrained [19], multi-objective constrained [20], bi-Fibonacci norm minimization constrained [55], [56], [57], nuclear norm with the local geometric regularized [58], smooth parafac decomposed [59], fast learning-based [12], and hypergraph regularized [66] ones.

Commonly, WSNs are deployed to the designed locations to collect sensing data with the preset time interval for a long time. As a result, spatiotemporal correlation is an inherent

characteristic of sensing data collected from WSNs [38], [39], [40], [46]. Hence, the spatiotemporal correlation has been frequently incorporated into an LRMA model to improve its recovery accuracy [36], [45]. For example, spatiotemporal correlation is usually captured as the neighborhood regularization constraint [38]. Besides, since many spatiotemporal signals exhibit graph structure, graph regularization is usually considered during building an LRMA model [23], [29]. For example, Mao and Gu [28] adopted a graph Laplacian matrix to favor smooth solutions with respect to the graph topology. Mao et al. [40] constructed a hybrid graph regularization to characterize the global and local smoothness of spatiotemporal signals. Luo et al. [63] modeled spatiotemporal correlation with structure regularization to improve the feature selection process's accuracy.

However, although the above LRMA-based models are different from each other in terms of model design or learning algorithms, they all adopt an L_2-norm-oriented loss function to minimize the errors between ground truths and estimations. As well known, L_2-norm is sensitive to outliers [41]. Therefore, their robustness cannot be guaranteed when there are outliers during data recovery in WSNs.

Compared with L_2-norm, L_1-norm is much more robust to outliers [64]. For example, the robust semi-supervised feature selection methods were achieved by incorporating L_1-norm into the feature selection processes [61], [62]. A robust spatial–temporal graph synchronous aggregation model is achieved for traffic prediction by adopting Huber loss [65]. Hence, L_1-norm has been widely employed to alleviate the adverse effects of outliers in industrial applications, such as recommender systems [26], [41], Web service [27], and image processing [47], [48], [49], [50]. Motivated by this, the proposed LFA-STSR model aggregates the L_1-norm into its loss part to significantly improve its robustness to outliers. Besides, the spatiotemporal correlation is also incorporated into the proposed LFA-STSR model as the regularization constraint. With such designs, the proposed LFA-STSR model possesses the merits of both high accuracy and robustness during data recovery, making it significantly outperform its peers.

III. Preliminaries

A. Problem of Data Recovery in WSNs

Adopted symbols are summarized in Nomenclature. Generally speaking, the collected data from WSNs are incomplete due to various reasons, e.g., sudden sensors breakdown, saving energy by deliberately shutting down partial nodes, human errors, uncontrollable factors, etc. [7]. The problem of data recovery in WSNs is to recover the missing data based on the observed data. Next, we formulate this problem.

Definition 1 (Problem of Data Recovery in WSNs): Given a matrix Y with M rows and N columns to record the data generated by M sensors during N time slots. Each entry $y_{i,j}$ of Y at ith row and jth column records the data generated by ith sensor at jth time slot, where $i \in \{1, 2, \ldots, M\}$ and $j \in \{1, 2, \ldots, N\}$. Let Y_K and Y_U denote the known and unknown entry sets of Y, respectively. The problem of data recovery in WSNs is to estimate the Y_U based on the Y_K only. The value of Y_K/Y_U is denoted as the sampling rate.

B. LFA Model

The LFA model is derived from matrix factorization techniques. It is commonly utilized for representing sparse matrices in recommender systems [26], [41] and Web service selection [27]. Given a sparse matrix, an LFA model training two low-dimensional latent feature matrices using the available observed data only [41]. Moving forward, we formulate the LFA model in the context of data recovery in WSNs.

Definition 2 (An LFA Model for Data Recovery in WSNs): To recover the unknown data Y_U of Y from WSNs, an LFA model is to train two latent feature matrices S and U to make Y's rank-k approximation by minimizing the differences between known and estimated entries on Y_K only, where the sizes of S and U are $M \times k$ and $N \times k$, respectively, $k \ll \min\{M, N\}$. The approximation of Y is achieved by $Y = SU^T$.

Currently, the existing LFA models adopt the L_2-norm to measure the differences between known and estimated entries on Y_K. Then, the objective $\varepsilon(S, U)$ of an LFA model in WSNs is

$$\varepsilon(S, U) = \left\| J \circ \left(Y - SU^T\right) \right\|_{L_2}^2 \tag{1}$$

where J is an $M \times N$ binary index matrix whose entry at ith row and jth column is 1 if $y_{i,j}$ is observed and is 0 if $y_{i,j}$ is not observed, $\circ$ denotes the Hadamard product to perform element-wise multiplication between two matrices, and $||\cdot||_{L2}$ denotes the L_2-norm of a matrix. To avoid overfitting on Y_K, Tikhonov regularization can be incorporated into (1) as follows [11], [41]:

$$\varepsilon(S, U) = \left\| J \circ \left(Y - SU^T\right) \right\|_{L_2}^2 + \lambda\left(\|S\|_F^2 + \|U\|_F^2\right) \tag{2}$$

where λ is a hyperparameter that controls the regularization intensity, and $||\cdot||_F$ denotes the Frobenius norm of a matrix.

IV. Proposed LFA-STSR Model

The proposed LFA-STSR model improves the basic LFA model for data recovery in WSNs from two aspects. First, it exploits the spatiotemporal correlation as the regularization constraint to improve recovery accuracy. Second, it employs the L_1-norm-oriented loss function to improve its robustness to outliers. The next is to introduce the proposed LFA-STSR model.

A. Exploiting Spatiotemporal Correlation

The spatiotemporal correlation of WSNs has two basic characteristics [21], [22], [23], [24], [25], i.e., spatial smoothness and temporal smoothness. Spatial smoothness indicates that data collected from geographically nearby sensors are similar to each other at a time point. Temporal smoothness denotes that the data collected from the same sensors vary smoothly over time.

1) Spatial Smoothness: To capture the spatial smoothness, an undirected weighted sensor connection graph $G = (V, E, W)$ is constructed, where V ($|V| = M$) is the set of vertices representing the sensors, E is the edge set in which each edge denotes an association between two sensors, and W is the weighted adjacency matrix in which each entry $w_{i,i'}$ at

ith row and i'th column denotes the quantitative relationship between ith and i'th vertices. Given V, the distance matrix F between vertices is calculated according to their coordinates, where each entry $f_{i,i'}$ at ith row and i'th column denotes the Euclidean distance between ith and i'th vertices. According to F, W can be obtained by selecting each vertex's k-nearest vertices as

$$w_{i,i'} = \begin{cases} \dfrac{1}{f_{i,i'}^2}, & \text{if } i' \in \text{Top}_k(i) \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where $\text{Top}_k(i)$ denotes the top-k nearest vertices set of the ith vertex. Let L denote the Laplacian matrix of graph G defined as

$$L = \begin{bmatrix} \text{sum}(w_{1,.}) & -w_{1,2} & \cdots & -w_{1,M-1} & -w_{1,M} \\ -w_{2,1} & \text{sum}(w_{2,.}) & \cdots & -w_{2,M-1} & -w_{2,M} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -w_{M,1} & -w_{M,2} & \cdots & -w_{M,M-1} & \text{sum}(w_{M,.}) \end{bmatrix}_{M\times M} \tag{4}$$

where $\text{sum}(w_{i,.})$, $i \in \{1, 2, \ldots, M\}$, denotes the sum of ith row of W matrix. Then, the spatial smoothness is captured by L.

2) Temporal Smoothness: The measurements obtained from a sensor, represented as a row within a matrix Y, exhibit temporal correlation with consistency and smoothness. In other words, two adjacent elements in a row of Y should have very similar values. Therefore, it is expected that the differences between two adjacent time slots in a row of the reconstructed matrix should be minimized. To this end, we construct the temporal differential matrix D as follows:

$$D = \begin{bmatrix} -1 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & -1 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}_{N\times N-1}. \tag{5}$$

By making $SU^T D$, the temporal smoothness of the reconstructed matrix is transferred to $[\hat{y}_{i,2} - \hat{y}_{i,1}, \hat{y}_{i,3} - \hat{y}_{i,2}, \ldots, \hat{y}_{i,N} - \hat{y}_{i,N-1}]$. Then, the temporal smoothness of the reconstructed matrix is guaranteed by minimizing $||SU^T D||2$ F as a regularization constraint.

3) Regularization Constraint of Spatiotemporal Correlation: After obtaining the Laplacian matrix L and the temporal differential matrix D, they are incorporated into (2) as the regularization constraints as

$$\varepsilon(S,U) = \left\| J \circ \left(Y - SU^T\right)\right\|_{L_2}^2 + \lambda\left(\|S\|_F^2 + \|U\|_F^2\right) + \underbrace{z_1\left\|\left(LSU^T\right)\right\|_F^2}_{\text{Spatial smoothness constraint}} + \underbrace{z_2\left\|SU^T D\right\|_F^2}_{\text{Temporal smoothness constraint}} \tag{6}$$

where z_1 and z_2 are two hyperparameters controlling the effects of spatial and temporal smoothness, respectively.

B. Improving the Robustness Based on the L_1-Norm-Oriented Loss

In (6), the differences between known and estimated entries on Y_K are measured by L_2-norm, which cannot guarantee robustness as L_2-norm is sensitive to outliers [41]. To address this issue, L_1-norm-oriented *Loss* is incorporated into (6) to improve its robustness as follows:

$$\varepsilon(S,U) = \underbrace{\beta_1\left\|J \circ \left(Y - SU^T\right)\right\|_{L_1}}_{L_1-\text{norm-oriented Loss}} + \underbrace{\beta_2\left\|J \circ \left(Y - SU^T\right)\right\|_{L_2}^2}_{L_2-\text{norm-oriented Loss}} + \lambda\left(\|S\|_F^2 + \|U\|_F^2\right) + z_1\left\|\left(LSU^T\right)\right\|_F^2 + z_2\left\|SU^T D\right\|_F^2 \tag{7}$$

where β_1 and β_2 are aggregation weights controlling the effects of L_1 and L_2 norms, respectively. To finely ensemble the advantages of both L_1 and L_2 norms, β_1 and β_2 are made self-adaptive according to their estimation errors. The main idea is to increase β_1 and decrease β_2 if L_1-norm's partial estimation errors are smaller than that of L_2-norm at tth training iteration and vice versa. The next is to formulate this strategy.

Definition 3: Let O_1^t and O_2^t be the partial estimation errors depending on L_1-norm and L_2-norm at the tth iteration, respectively. Then, O_1^t and O_2^t are computed as follows:

$$O_1^t = \left\| J \circ \left(Y - S^t\left(U^t\right)^T\right)\right\|_{L_1}$$
$$O_2^t = \left\| J \circ \left(Y - S^t\left(U^t\right)^T\right)\right\|_{L_2}^2. \tag{8}$$

where S^t and U^t denote the states of S and U at the tth iteration.

Definition 4: Let B_{O1}^t and B_{O2}^t be the cumulative estimation errors corresponding to O_1^t and O_2^t until the tth iteration, respectively. Then, B_{O1}^t and B_{O2}^t are computed as follows:

$$B_{O_1}^t = \sum_{\omega=1}^{t} O_1^{\omega}, \quad B_{O_2}^t = \sum_{\omega=1}^{t} O_2^{\omega}. \tag{9}$$

Based on definitions 3 and 4, β_1 and β_2 can be set self-adaptively during the training process as follows:

$$\beta_1^t = \frac{e^{-\sigma B_{O_1}^{t-1}}}{e^{-\sigma B_{O_1}^{t-1}} + e^{-\sigma B_{O_2}^{t-1}}}, \quad \beta_2^t = \frac{e^{-\sigma B_{O_2}^{t-1}}}{e^{-\sigma B_{O_1}^{t-1}} + e^{-\sigma B_{O_2}^{t-1}}} \tag{10}$$

where β_1^t and β_2^t denote the states of β_1 and β_2 at the tth iteration, and σ is a balance coefficient controlling the ensemble of L_1-norm and L_2-norm. σ can be set as $\sigma = (1/\ln T)^{1/2}$ where T is the maximum number of iterations.

C. Model Optimization

Notably, since the L_1-norm-oriented *Loss* term of (7) is not differentiable at the zero point, extending (7) into the single-element-oriented form as follows [11], [27], [41]:

$$\begin{aligned}\varepsilon(S,U) &= \underbrace{\beta_1 \sum_{y_{i,j}\in Y_K} \left|y_{i,j}-s_{i,.}u_{j,.}\right|}_{L_1-\text{norm-oriented Loss}} + \underbrace{\beta_2 \sum_{y_{i,j}\in Y_K} \left(y_{i,j}-s_{i,.}u_{j,.}\right)^2}_{L_2-\text{norm-oriented Loss}} \\ &+ \underbrace{z_1 \sum_{y_{i,j}\in Y_K} \left(\left(LSU^T\right)_{(i,j)}\right)^2}_{\text{Spatial smoothness constraint}} + \underbrace{z_2 \sum_{y_{i,j}\in Y_K} \left(\left(SU^TD\right)_{(i,j)}\right)^2}_{\text{Temporal smoothness constraint}} \\ &+ \underbrace{\lambda \sum_{y_{i,j}\in Y_K} \left(\left(s_{i,.}\right)^2+\left(u_{j,.}\right)^2\right)}_{\text{Tikhonov regularization}} \end{aligned} \tag{11}$$

where $s_{i,.}$ and $u_{j,.}$ denote the ith and the jth row vector of S and U, respectively. Then, considering the instant loss $\varepsilon_{i,j}$ of $\varepsilon(S,U)$ on a single entry $y_{i,j}$, then

$$\begin{aligned}\varepsilon_{i,j} &= \beta_1\left|y_{i,j}-s_{i,.}u_{j,.}\right| + \beta_2\left(y_{i,j}-s_{i,.}u_{j,.}\right)^2 + z_1\left(\left(LSU^T\right)_{(i,j)}\right)^2 \\ &+ z_2\left(\left(SU^TD\right)_{(i,j)}\right)^2 + \lambda\left(\left(s_{i,.}\right)^2+\left(u_{j,.}\right)^2\right). \end{aligned} \tag{12}$$

The optimization of (12) with respect to $s_{i,.}$ and $u_{j,.}$ can be achieved by the stochastic gradient descent (SGD) algorithm. Then, at the tth iteration, employing SGD to minimize (12) as follows:

$$\begin{cases} s_{i,.}^t = s_{i,.}^{t-1} - \eta\dfrac{\partial\varepsilon_{i,j}^{t-1}}{\partial s_{i,.}^{t-1}} \\ u_{j,.}^t = u_{j,.}^{t-1} - \eta\dfrac{\partial\varepsilon_{i,j}^{t-1}}{\partial u_{j,.}^{t-1}} \end{cases} \tag{13}$$

where $s^{t-1}i,.$, $u_{j,.}^{t-1}$, and $\varepsilon_{i,j}^{t-1}$ denote the states of $s_{i,.}$, $u_{j,.}$, and $\varepsilon_{i,j}$ at the $(t-1)$th iteration, and η denotes the learning rate of SGD. Let $\Delta_{i,j}^{t-1} = y_{i,j} - s_{i,.}^{t-1}u_{j,.}^{t-1}$ be the estimation error on a single entry $y_{i,j}$ at the $(t-1)$th iteration. By combining (12) into (13), the updating rules of $s^ti,.$ and $u_{j_*}^t$ on a single entry $y_{i,j}$ at the tth iteration are obtained as follows:

$$\Delta_{i,j}^{t-1} > 0: \begin{cases} s_{i,.}^t = s_{i,.}^{t-1} + \beta_1^{t-1}\eta u_{j,.}^{t-1} + 2\beta_2^{t-1}\eta u_{j,.}^{t-1}\Delta_{i,j}^{t-1} \\ \quad - 2z_1\eta\left(\left(LL^T\right)_{i,.}S^{t-1}\left(u_{j,.}^{t-1}\right)^T\right)u_{j,.}^{t-1} \\ \quad - 2z_2\eta\left(s_{i,.}^{t-1}\left(U^{t-1}\right)^T\left(DD^T\right)_{.,j}\right)u_{j,.}^{t-1} \\ \quad - 2\eta\lambda s_{i,.}^{t-1} \\ u_{j,.}^t = u_{j,.}^{t-1} + \beta_1^{t-1}\eta s_{i,.}^{t-1} + 2\beta_2^{t-1}\eta s_{i,.}^{t-1}\Delta_{i,j}^{t-1} \\ \quad - 2z_1\eta\left(\left(LL^T\right)_{i,.}S^{t-1}\left(u_{j,.}^{t-1}\right)^T\right)s_{i,.}^{t-1} \\ \quad - 2z_2\eta\left(s_{i,.}^{t-1}\left(U^{t-1}\right)^T\left(DD^T\right)_{.,j}\right)s_{i,.}^{t-1} \\ \quad - 2\eta\lambda u_{j,.}^{t-1}. \end{cases} \tag{14}$$

D. Theoretical Analysis

1) Ensemble of L_1-Norm- and L_2-Norm-Oriented Loss:

Definition 5: Let $O_{1,2}^t$ be the partial estimation errors depending on both L_1-norm- and L_2-norm-oriented *Loss* at the tth iteration. Then, $O_{1,2}^t$ is computed as follows:

$$\begin{aligned} O_{1,2}^t &= \beta_1^t\left\|J\circ\left(Y-S^t\left(U^t\right)^T\right)\right\|_{L_1} \\ &\quad + \beta_2^t\left\|J\circ\left(Y-S^t\left(U^t\right)^T\right)\right\|_{L_2}^2 \\ &= \beta_1^tO_1^t + \beta_2^tO_2^t. \end{aligned} \tag{15}$$

Definition 6: Let $B_{O1,2}^t$ be the cumulative estimation errors corresponding to $O_1^t, 2$ until the tth iteration. $B_{O1,2}^t$ is computed as

$$B_{O_{1,2}}^t = \sum_{\omega=1}^{t} O_{1,2}^{\omega}. \tag{16}$$

Theorem 1: Assuming that O_1^t and O_2^t lie in the scale of $[0,1]$. If β_1^t and β_2^t are set as (10) during the training processes, then the following inequality holds:

$$B_{O_{1,2}}^T \le \min\left\{B_{O_1}^T, B_{O_2}^T\right\} + \frac{\ln 2}{\sigma} + \frac{\sigma T}{8}. \tag{17}$$

Proof: Let us recall the Hoeffding inequality [51].

Lemma 1: Let Q be a random variable fulfilling $a \le Q \le b$, $\forall l \in \mathbb{R}$, the following inequality holds:

$$\ln E\left[e^{lQ}\right] \le lEQ + \frac{l^2(b-a)^2}{8}. \tag{18}$$

The detailed proof of Lemma 1 is given in [33]. Note that B_{O1}^T and B_{O2}^T are bounded by

$$F_t = e^{-\sigma B_{O_1}^t} + e^{-\sigma B_{O_2}^t} \tag{19}$$

when $t=0$, $F_0=2$. Based on (19), having the following inference:

$$\ln\frac{F_T}{F_0} \ge -\sigma\min\left\{B_{L_1}^T, B_{L_2}^T\right\} - \ln 2. \tag{20}$$

Besides, $\forall t \in \{1,\ldots,T\}$, based on (10) and (19), then

$$\ln\frac{F_t}{F_{t-1}} = \ln\left(\beta_1^te^{-\sigma O_1^t} + \beta_2^te^{-\sigma O_2^t}\right). \tag{21}$$

Let $l = -\sigma$ and $Q \in \{O_1^t, O_2^t\}$ with O_1^t and O_2^t in the scale of [0, 1]. Note that β_1^t and β_2^t are the probabilities for O_1^t and O_2^t, respectively. Thus, having

$$l\mathbb{E}Q = -\sigma\left(\beta_1^t O_1^t + \beta_2^t O_2^t\right). \tag{22}$$

Then based on Lemma 1, (15), (21), and (22), having

$$\ln\frac{F_t}{F_{t-1}} \le -\sigma\left(\beta_1^t O_1^t + \beta_2^t O_2^t\right) + \frac{\sigma^2}{8} = -\sigma O_{12}^t + \frac{\sigma^2}{8}. \tag{23}$$

Considering the accumulation of (23) as t increases from 1 to T to obtain

$$\ln\frac{F_T}{F_0} \le -\sigma B_{L_{12}}^T + \frac{\sigma^2}{8}T. \tag{24}$$

By combining (20) and (24) to obtain

$$B_{O_{1,2}}^T \le \min\left\{B_{O_1}^T, B_{O_2}^T\right\} + \frac{\ln 2}{\sigma} + \frac{\sigma T}{8}. \tag{25}$$

Therefore, Theorem 1 holds. □

In Theorem 1, by setting $\sigma = (1/\ln T)^{1/2}$, the upper bound becomes $\min\{B_{O1}^T, B_{O2}^T\} = \ln 2\sqrt{\ln T} + T/(8\sqrt{\ln T})$, where the term of $\ln 2\sqrt{\ln T} + T/(8\sqrt{\ln T})$ is linearly bounded by the number of iterations. Then, the following proposition is obtained.

Proposition 1: By setting $\sigma = (1/\ln T)^{1/2}$, the following inequality holds:

$$\begin{cases} B_{O_{1,2}}^T \le B_{O_2}^T + \text{const} < B_{O_1}^T + \text{const}, & \text{if } B_{O_1}^T > B_{O_2}^T \\ B_{O_{1,2}}^T \le B_{O_1}^T + \text{const} < B_{O_2}^T + \text{const}, & \text{otherwise} \end{cases} \tag{26}$$

where $\lim_{T\to\infty}\text{const} = 19.45$.

Proof: Let us recall the Bernstein inequality [33].

Lemma 2: Let Q be a random variable from 0 to 1, and $\forall l \in \mathbb{R}$, the following inequality holds [33]:

$$\ln\mathbb{E}\left[e^{lQ}\right] \le \left(e^l - 1\right)\mathbb{E}Q. \tag{27}$$

When $B_{O1}^T > B_{O2}^T$, (20) is reduced into

$$\ln\frac{F_T}{F_0} \ge -\sigma B_{O_2}^T - \ln 2. \tag{28}$$

By combining Lemma 2, (15), (21), and (22), the following inequality is achieved:

$$\ln\frac{F_t}{F_{t-1}} \le \left(e^{-\sigma} - 1\right)O_{1,2}^t. \tag{29}$$

Considering the accumulation of (29) according to (26) as t increases from 1 to T to have the following inferences:

$$\ln\frac{F_T}{F_0} \le \left(e^{-\sigma} - 1\right)B_{O_{1,2}}^T. \tag{30}$$

By combining (28) and (30) to obtain

$$B_{O_{1,2}}^T \le \frac{\sigma}{1 - e^{-\sigma}}B_{L_2}^T + \frac{\ln 2}{1 - e^{-\sigma}}. \tag{31}$$

Let $\psi = \sigma/(1 - e^{-\sigma})$ and $\Phi = \ln 2/(1 - e^{-\sigma})$. With $\sigma = (1/\ln T)^{1/2}$, to obtain $\lim_{T\to\infty}\psi = 1$ following L'Hopital's Rule. Moreover, with the partial derivative of Φ with T, having the inference that $\lim_{T\to\infty}(\mathrm{d}\Phi/\mathrm{d}T) \to 0$. Thus, Φ becomes a constant (i.e., 19.45) when $T \to \infty$. Finally, (31) can be reduced to

$$B_{O_{1,2}}^T \le \left[B_{O_2}^T\right]^+ + [19.45]^- \le B_{O_2}^T + \underset{<19.45}{\text{const}}. \tag{32}$$

Hence, we obtain that $B_{O1}^T,2 \le B_{O2}^T + \text{const} < B_{O1}^T + \text{const}$ under the condition of setting $\sigma = (1/\ln T)^{1/2}$ when $B_{O1}^T > B_{O2}^T$ and $B_{O1}^T,2 \le B_{O1}^T + \text{const} < B_{O2}^T + \text{const}$ when $B_{O1}^T < B_{O2}^T$. Therefore, Proposition 1 holds. □

Remark 1: Proposition 1 states that $B_{O1}^T,2$ is bounded by $\min\{B_{O1}^T, B_{O2}^T\} + \text{const}$ with the condition of $\sigma = (1/\ln T)^{1/2}$, which indicates that the self-adaptive β_1^t and β_2^t set by (10) guarantee that $B_{O1}^T,2$ is always comparable to, or not larger than, both B_{O1}^T and B_{O2}^T during the training process. Therefore, the LFA-STSR model possesses the fine ensemble effects of L_1-norm-oriented *Loss*'s robustness and L_2-norm-oriented *Loss*'s stability to precisely estimate the missing data Y_U based on the Y_K only.

2) Recovery Error Bound:

Proposition 2: For any $H \in \mathbb{R}^{M\times N}$ with a constant $\alpha \in [0, 1)$, if the following inequality holds:

$$\left\|H - \beta_1 J \circ \text{sign}(H) - \beta_2 J \circ H - z_1 L^T L H - z_2 H D D^T\right\|_F \le \alpha\|H\|_F \tag{33}$$

where sign$(\cdot)$ denotes the sign function. Then, the recovery error bound of the proposed LFA-STSR model is

$$\|X - Y\|_F \le \frac{\lambda\sqrt{2\text{rank}(Y)} + z_1\left\|L^T L Y\right\|_F + z_2\left\|Y D D^T\right\|_F}{1 - \alpha} \tag{34}$$

where X is the recovery matrix obtained by $X = SU^T$.

Proof: First, let us recall [54, eq. (6)] as follows:

$$\begin{aligned}\text{rank}(X) = \|X\|_* &= \min_{X=SU^T}\|S\|_F\|U\|_F \\ &= \min_{X=SU^T}\frac{1}{2}\left(\|S\|_F^2 + \|U\|_F^2\right)\end{aligned} \tag{35}$$

then, we upper bound the recovery error of the proposed LFA-STSR model as follows:

$$\begin{aligned}\|X - Y\|_F \le &\underbrace{\left\|X - \text{SVD}_\lambda\left(Y - z_1 L^T L Y - z_2 Y D D^T\right)\right\|_F}_{\text{A}} \\ &+ \underbrace{\left\|\text{SVD}_\lambda\left(Y - z_1 L^T L Y - z_2 Y D D^T\right) - Y\right\|_F}_{B}.\end{aligned} \tag{36}$$

Next, we upper bound the terms A and B of (36), respectively. First, by Karush-Kuhn-Tucker (KKT) condition to have

$$\begin{aligned}X = Z, \quad &\beta_1 J \circ \text{sign}(X - Y) + \beta_2 J \circ (X - Y) + z_1 L^T L X \\ &+ z_2 X D D^T = 0, \quad 0 \in \lambda\partial_{\|\cdot\|_*}(Z).\end{aligned} \tag{37}$$

Based on the three conditions of (37), having

$$\begin{aligned}\left(-\beta_1 J \circ \text{sign}(X - Y) - \beta_2 J \circ (X - Y) - z_1 L^T L X\right. \\ \left. - z_2 X D D^T\right) \in \lambda\partial_{\|\cdot\|_*}(X).\end{aligned} \tag{38}$$

Then, transforming (38) as follows:

$$\begin{aligned}\left(X - \beta_1 J \circ \text{sign}(X - Y) - \beta_2 J \circ (X - Y) - z_1 L^T L X\right. \\ \left. - z_2 X D D^T\right) \in \left(\lambda\partial_{\|\cdot\|_*}(X) + X\right).\end{aligned} \tag{39}$$

According to the proof of [53, Th. 3], i.e., for any $r \in \mathbb{R}^+$, having $\mathrm{SVD}_r = (\mathbf{I} + r\lambda\partial_{\|\cdot\|_*})^{-1}$. Based on (39) and make $r = \lambda$, having

$$X = \mathrm{SVD}_\lambda\begin{pmatrix} X - \beta_1 J \circ \mathrm{sign}(X - Y) \\ -\beta_2 J \circ (X - Y) - z_1 L^T L X - z_2 X D D^T \end{pmatrix}. \tag{40}$$

Let us recall [53, Lemma 1], i.e., for any two matrices H_1 and H_2, and any $r \in \mathbb{R}^+$, having

$$\|\mathrm{SVD}_\lambda(H_1) - \mathrm{SVD}_\lambda(H_2)\|_F \le \|H_1 - H_2\|_F. \tag{41}$$

Then, based on (40), (41), and term A of (36), having

$$A \le \left\| \begin{matrix} X - Y - \beta_1 J \circ \mathrm{sign}(X - Y) - \beta_2 J \circ (X - Y) \\ -z_1 L^T L (X - Y) - z_2 (X - Y) D D^T \end{matrix} \right\|_F. \tag{42}$$

By making $H = X - Y$, according to (33) to have

$$A \le \alpha \|H\|_F. \tag{43}$$

Then, by considering the term B of (36) to have

$$\begin{aligned} B &\le \left\|\mathrm{SVD}_\lambda\left(Y - z_1 L^T L Y - z_2 Y D D^T\right)\right. \\ &\quad \left. - Y - z_1 L^T L Y - z_2 Y D D^T \right\|_F \\ &\quad + z_1 \left\| L^T L Y \right\|_F + z_2 \left\| Y D D^T \right\|_F \\ &\le \|\Lambda_\lambda(\Sigma) - (\Sigma)\|_F + z_1 \left\| L^T L Y \right\|_F + z_2 \left\| Y D D^T \right\|_F \\ &\le \lambda\sqrt{2\mathrm{rank}(Y)} + z_1 \left\| L^T L Y \right\|_F + z_2 \left\| Y D D^T \right\|_F \end{aligned} \tag{44}$$

where $\Lambda_\lambda(\cdot)$ denotes the function of $\Lambda_\lambda(\cdot) := \mathrm{sign}(\cdot)\max(|\cdot| - \lambda, 0)$ and Σ is to collect the singular values of $(Y - z_1 L^T L Y - z_2 Y D D^T)$ as its diagonal elements. By substituting (43) and (44) into (36) to obtain (34). Therefore, Proposition 2 holds. □

E. Algorithm Design and Time Complexity

In (14), the spatiotemporal correlation has much computational burden. Considering the spatiotemporal correlation by every few entries to improve computational efficiency. To this end, we first remove the spatiotemporal correlation from (14) as follows:

$$\begin{aligned} &\Delta_{i,j}^{t-1} \\ &> 0: \begin{cases} s_{i,\cdot}^t = s_{i,\cdot}^{t-1} + \beta_1^{t-1}\eta u_{j,\cdot}^{t-1} + 2\beta_2^{t-1}\eta u_{j,\cdot}^{t-1}\Delta_{i,j}^{t-1} \\ \qquad - 2\eta\lambda s_{i,\cdot}^{t-1} \\ u_{j,\cdot}^t = u_{j,\cdot}^{t-1} + \beta_1^{t-1}\eta s_{i,\cdot}^{t-1} + 2\beta_2^{t-1}\eta s_{i,\cdot}^{t-1}\Delta_{i,j}^{t-1} \\ \qquad - 2\eta\lambda u_{j,\cdot}^{t-1} \end{cases} \\ &\Delta_{i,j}^{t-1} \\ &< 0: \begin{cases} s_{i,\cdot}^t = s_{i,\cdot}^{t-1} - \beta_1^{t-1}\eta u_{j,\cdot}^{t-1} + 2\beta_2^{t-1}\eta u_{j,\cdot}^{t-1}\Delta_{i,j}^{t-1} \\ \qquad - 2\eta\lambda s_{i,\cdot}^{t-1} \\ u_{j,\cdot}^t = u_{j,\cdot}^{t-1} - \beta_1^{t-1}\eta s_{i,\cdot}^{t-1} + 2\beta_2^{t-1}\eta s_{i,\cdot}^{t-1}\Delta_{i,j}^{t-1} \\ \qquad - 2\eta\lambda u_{j,\cdot}^{t-1}. \end{cases} \end{aligned} \tag{45}$$

Then, we adopt an index μ to control the incorporation of the spatiotemporal correlation. As such, Algorithm LFA-STSR is designed as shown in the Algorithm LFA-STSR. Its time complexity is mainly decided by the maximum number of iterations, the number of target sparse matrix's known entries, the number of computing spatiotemporal correlations, and the latent feature dimension. Thus, its time cost is $\Theta(T \times |Y_K| \times k \times M \times N/\mu)$.

Algorithm 1 Algorithm LFA-STSR

Input: L, Y_K; **Output:** S, U.

```
initialize k, λ, η, β1 = 0.5, β2 = 0.5, T, μ
initialize S randomly
initialize U randomly
while t ≤ T && not converge
 for j = 1 to N // input data by column of Y
  Index=0
  for each y_{i,j} in Y_K
   if Index≠ μ
     update s^t_{i,.}, . according to (45)
    else update s^t_{i,.} according to (14)
   end if
  end for
  Index= Index+1
 end for
 for i = 1 to M // input data by row of Y
   Index=0
    for each y_{i,j} in Y_K
  if Index≠ μ
     update u^t_{j,.} according to (45)
   else update u^t_{j,..} according to (14)
  end if
 end for
  Index= Index+1
end for
 update β^t_1 and β^t_2 according to (8)–(10)
 t = t + 1
 end while
```

V. Experiments

The experiments are designed to answer the following research questions (RQs).

RQ. 1. Does the proposed LFA-STSR model outperform the related state-of-the-art models in recovering missing data for WSNs?

RQ. 2. How do outlier data influence the performance of the involved recovery models?

RQ. 3. How do the L_1 and L_2 norms influence the performance of the proposed LFA-STSR model?

RQ. 4. How do hyperparameters influence the performance of the proposed LFA-STSR model?

A. General Settings

1) Datasets: Four benchmark datasets are selected to conduct the experiments. They are real datasets generated by WSNs, including PM2.5, CO, and sea surface temperature. Table I summarizes their properties.

2) Evaluation Metrics: Mean absolute error (MAE) and root mean-squared error (RMSE) are two common evaluation metrics used to assess the accuracy of a predictive model or algorithm [18], [19], [20]. Both MAE and RMSE provide

TABLE I
PROPERTIES OF EXPERIMENTAL DATASETS

No.	Name	$\|M\|$	$\|N\|$	Time	Minimum/ Maximum
D1	Beijing PM2.5[2] concentration [36]	35	8647	2014-05–2015-04	3.0/773.7
D2	Beijing CO Concentration [2] [37]	35	8647	2014-05–2015-04	0.1/20.0
D3	Sea surface Temperature [3] [34]	70	1733	1870-01–2014-12	0.01/30.31
D4	Daily mean CO concentration[4] [35]	74	365	2010-01–2010-12	0.1/2.9

TABLE II
SUMMARY OF COMPARED MODELS

Model	Description
ST-LRMA [38]	It jointly applies the global and local correlations to build an LRMA model for recovering the spatio-temporal signal.
BR-TVGS [39]	It is a batch reconstruction-based LRMA model for recovering time-varying graph signals.
LRDS [40]	It introduces the differential smooth prior of time-varying graph signals into the LRMA model for recovering the spatio-temporal signal.
RRImpu [42]	It leverages optimal transport distances to impute missing data values. It is an end-to-end machine learning method.
L^3F [41]	It is an L_1-and-L_2-norm-oriented LFA model. It does not consider the spatio-temporal correlation.
TRSS [60]	It is a time-varying graph signal reconstruction model via sobolev smoothness algorithm.
LFA-STSR	It is the latent feature analysis (LFA) based spatio-temporal signal recovery (STSR) model proposed in this paper.

a quantitative measure of a model's prediction performance. RMSE is commonly used when larger errors are more critical due to the squared term, while MAE is often preferred when all errors are considered equally important due to the absolute term [38], [39], [40], [41], [42]. This article aims at recovering the missing data collected from WSNs, where outliers are mixed inevitably. Hence, RMSE can effectively reflect the recovery accuracy with many outliers, while MAE can effectively evaluate the recovery accuracy on ground-truth data. MAE and RMSE are calculated as follows:

$$\text{MAE} = \left(\sum_{y_{i,j}\in\Gamma} \left|y_{i,j} - \hat{y}_{i,j}\right|\right) \Big/ |\Gamma|$$

$$\text{RMSE} = \sqrt{\left(\sum_{y_{i,j}\in\Gamma} \left(y_{i,j} - \hat{y}_{i,j}\right)^2\right) \Big/ |\Gamma|}$$

where $\hat{y}_{i,j}$ denote the estimation of $y_{i,j}$ and Γ denotes the testing set, a lower value of RMSE/MAE indicates better accuracy.

3) Baselines: The proposed LFA-STSR model is compared with six related state-of-the-art models. Table II briefly describes these competitors.

4) Experimental Details: To simulate the missing data situation in WSNs, partial data are randomly sampled from the complete dataset as the training set, while the remaining data serve as the testing set. In the training set, half of the data is allocated for training the model, while the other half is used to validate performance and tune hyperparameters. Once the optimal hyperparameters are determined, the model is retrained using all the data from the training set. The maximum number of iterations is set to 1500. Additionally, the training of a model will be terminated if the error difference between two consecutive iterations becomes smaller than 10^{-6}. For each dataset, the spatial correlation of the observation sites is characterized using the K-nearest neighbor method, selecting five nodes with the smallest distance. Each test is repeated five times, and the average results are reported. All experiments are conducted on a PC equipped with a 3.4-GHz i7 CPU and 64-GB RAM.

B. Performance Comparison (RQ. 1)

This set of experiments increases the sampling rate from 0.1 to 0.9 for comparing the performance of all the involved models. The comparison results on all the datasets are recorded in Tables III–VI. To better under these results, the statistical analyses of win/loss, Wilcoxon signed-ranks test [54], and Friedman test [54] are conducted. The win/loss is to count how many cases that LFA-STSR has lower/higher MAE and RMSE than each comparison model as the sampling rate increases from 0.1 to 0.9, respectively. The Wilcoxon signed-ranks test is a nonparametric pairwise comparison method to check whether LFA-STSR has significantly better recovery accuracy than each comparison model by the significance level of p-value.

The Friedman test is to compare the performance of multiple models on multiple cases simultaneously by the F-rank value. A lower F-rank value indicates a higher recovery accuracy. Tables III–VI shows that the MAE/RMSE decreases as the sampling rate increases. Besides, we have the following important findings.

1) LFA-STSR achieves the lowest RMSE/MAE in most cases. It shows a lower performance compared to time-varying reconstruction via sobolev smoothness (TRSS) on D1, TRSS on D2, and low rank and differential smoothness (LRDS) on D4, with only 4, 5, and 4 cases, respectively.
2) All the p-values on four datasets are smaller than 0.05, which verifies that LFA-STSR possesses significantly higher recovery accuracy than all the comparison models with a significance level of 0.05 on the tested datasets.
3) LFA-STSR achieves the lowest F-rank value among all the models, further verifying that it has the highest recovery accuracy among all the models.

Therefore, the experimental results demonstrate that LFA-STSR has significantly higher accuracy than the six comparison models in recovering the missing data from WSNs.

C. Influence of Outlier Data (RQ. 2)

This set of experiments tests the influence of outlier data to all the involved models. The sampling rate is set as 0.1. The specific method of adding outlier data is as follows.

1) Randomly selecting an unknown entry as the outlier entry for the sparse matrix Y from WSNs.

TABLE III
COMPARISON RESULTS OF RECOVERY ACCURACY ON D1 WITH DIFFERENT SAMPLING RATES

Sampling rate	Metric	ST-LRMA	BR-TVGS	LRDS	RRImpu	L^{3}F	TRSS	LFA-STSR
0.1	MAE	19.1256±1.2976	23.6990±0.6210	21.3191±1.2525	26.5679±1.6661	23.7305±1.1598	18.4799±1.2521	17.4034±1.6710
	RMSE	32.6456±1.4268	40.1144±0.6417	38.0017±1.5285	43.6397±1.8321	43.0706±1.3881	31.0868±1.3368	29.6706±1.6076
0.2	MAE	16.7820±1.2693	15.5888±0.6529	15.4546±0.9856	22.7131±1.1832	18.7451±1.0670	15.3071±1.0448	14.5201±1.7143
	RMSE	29.6656±1.3816	29.8892±0.7345	27.1822±1.0275	38.6820±1.6750	33.5317±1.7307	27.2937±1.7222	25.3580±1.0093
0.3	MAE	15.4942±1.1076	14.7988±0.9809	13.5559±0.6900	21.9314±1.6672	17.2958±1.0369	13.0247±1.0948●	13.1528±1.0848
	RMSE	27.8261±0.9092	25.9836±1.3401	24.0348±1.0233	37.1785±1.1196	30.6239±1.3698	23.9525±1.1143	23.0036±1.5482
0.4	MAE	14.5286±1.6449	13.1289±0.7523	12.3299±1.0472	21.8399±1.9836	16.2178±1.1797	11.6659±1.9821●	12.1801±1.0880
	RMSE	26.1977±1.3189	23.0221±1.0581	21.7812±1.3860	36.5695±1.7551	28.6907±1.3281	21.2653±1.0638	21.2089±1.5978
0.5	MAE	13.8788±1.2694	11.9313±1.0566	11.5769±1.0697	22.0211±1.3289	15.4683±1.7812	10.4380±1.073	10.0774±1.4512
	RMSE	25.5460±1.0071	21.6123±1.0760	20.9354±0.6273	37.0273±1.8378	27.7926±1.5310	19.7193±0.7059	18.8695±1.6227
0.6	MAE	13.3955±1.4734	11.2388±0.6120	10.8765±1.1926	22.1146±1.4874	14.5673±1.0942	9.6843±1.3738●	9.8194±1.1818
	RMSE	24.4877±1.3903	20.4831±0.7051	20.0635±0.9832	37.4808±1.9235	26.0242±1.0852	18.3888±1.6211	18.3317±1.0892
0.7	MAE	12.7804±1.5222	10.5210±0.9123	10.1808±1.2246	23.9995±1.4544	14.3649±1.1452	9.0596±1.4828	8.8452±1.0991
	RMSE	23.3757±1.6290	19.7710±1.0732	19.0313±1.5264	39.5898±1.9725	25.9349±1.3777	17.2383±1.0452	16.6990±1.4824
0.8	MAE	12.3630±0.9456	9.8272±0.6785	9.5690±0.7054	23.2691±1.0948	13.8012±1.1559	8.2602±1.5085●	8.2616±1.5350
	RMSE	22.7555±1.9798	18.7477±1.0086	18.0317±1.0676	38.8918±1.5065	25.3508±1.0486	16.3191±1.6447	16.0696±1.6489
0.9	MAE	11.7224±1.1232	9.2217±1.4675	8.9917±0.5017	24.1504±1.0819	13.5331±1.4572	8.1238±1.4842	7.7682±1.4846
	RMSE	21.2764±1.5991	16.7541±0.9375	16.4098±1.0589	39.1380±1.0891	24.3531±1.2362	15.0165±1.0525	14.5509±1.6721
Statistical Analysis	win/loss♦	18/0	18/0	18/0	18/0	18/0	14/4	Total 104/4
	p-value⋆	**0.0001**	**0.0001**	**0.0001**	**0.0001**	**0.0001**	**0.0027**	—
	F-rank*	4.72	4.17	3.06	7.00	6.00	1.83	**1.22**

● The case that LFA-STSR loses the comparison compared with other models; ♦ The win/loss count of LFA-STSR has lower/higher MAE and RMSE than each comparison model; ⋆ The *p*-value of the Wilcoxon signed-ranks test, where the accepted hypothese with a significance level of 0.05 are highlighted; * The F-rank of the Friedman test, where a smaller F-rank value denotes a higher recovery accuracy.

TABLE IV
COMPARISON RESULTS OF RECOVERY ACCURACY ON D2 WITH DIFFERENT SAMPLING RATES

Sampling rate	Metric	ST-LRMA	BR-TVGS	LRDS	RRImpu	L^{3}F	TRSS	LFA-STSR
0.1	MAE	0.3378±0.0411	0.4107±0.0361	0.3959±0.0261	0.4184±0.0474	0.3968±0.0263	0.3473±0.0226	0.3061±0.0327
	RMSE	0.6500±0.0342	0.7283±0.0384	0.7260±0.0441	0.7602±0.0455	0.7582±0.0361	0.6489±0.0417	0.5689±0.0328
0.2	MAE	0.3201±0.0270	0.3083±0.0164	0.3003±0.0125	0.3583±0.0244	0.3231±0.0125	0.2857±0.0183	0.2614±0.0225
	RMSE	0.6002±0.0173	0.5686±0.0189	0.5637±0.0208	0.6595±0.0209	0.6279±0.0202	0.5717±0.0289	0.4975±0.0243
0.3	MAE	0.2574±0.0160	0.2613±0.0289	0.2562±0.0189	0.3358±0.0198	0.2972±0.0146	0.2421±0.0126	0.2377±0.0157
	RMSE	0.5312±0.0146	0.4949±0.0192	0.4923±0.0226	0.6228±0.0213	0.5844±0.0135	0.4970±0.0135	0.4652±0.0162
0.4	MAE	0.2394±0.0213	0.2302±0.0189	0.2270±0.0149	0.3122±0.0106	0.2830±0.0134	0.2177±0.0191	0.2056±0.0149
	RMSE	0.5013±0.0232	0.4657±0.0184	0.4531±0.0117	0.5854±0.0174	0.5525±0.0108	0.4528±0.0148	0.4242±0.0153
0.5	MAE	0.2248±0.0217	0.2061±0.0163	0.2058±0.0123	0.3068±0.0213	0.2740±0.0104	0.1861±0.0132●	0.1917±0.0187
	RMSE	0.4849±0.0228	0.4305±0.0144	0.4231±0.0169	0.5803±0.0134	0.5455±0.0226	0.4147±0.0218	0.4123±0.0184
0.6	MAE	0.2153±0.0401	0.1905±0.0108	0.1892±0.0146	0.2983±0.0285	0.2607±0.0134	0.1712±0.0263●	0.1804±0.0151
	RMSE	0.4657±0.0377	0.3958±0.0173	0.3909±0.0197	0.5811±0.0131	0.5259±0.0188	0.3681±0.0287●	0.3840±0.0245
0.7	MAE	0.2013±0.0232	0.1750±0.0153	0.1741±0.0177	0.2991±0.0196	0.2500±0.0124	0.1570±0.0266●	0.1679±0.0171
	RMSE	0.4309±0.0245	0.3713±0.0177	0.3669±0.0152	0.5364±0.0193	0.5242±0.0127	0.3555±0.0292	0.3497±0.0137
0.8	MAE	0.1946±0.0228	0.1658±0.0189	0.1651±0.0115	0.3027±0.0235	0.2491±0.0186	0.1454±0.0224	0.1445±0.0207
	RMSE	0.4366±0.0208	0.3777±0.0142	0.3731±0.0102	0.5507±0.0159	0.5180±0.0146	0.3439±0.0273	0.3390±0.0218
0.9	MAE	0.1848±0.0196	0.1535±0.0133	0.1525±0.0206	0.3098±0.0191	0.2428±0.0141	0.1346±0.0229●	0.1358±0.0133
	RMSE	0.4232±0.0131	0.3522±0.0189	0.3502±0.0328	0.5787±0.0249	0.5113±0.0161	0.3307±0.0255	0.3240±0.0180
Statistical Analysis	win/loss	18/0	18/0	18/0	18/0	18/0	13/5	Total 103/5
	p-value	**0.0001**	**0.0001**	**0.0001**	**0.0001**	**0.0001**	**0.0226**	—
	F-rank	4.67	4.11	3.00	7.00	5.94	2.00	**1.28**

2) Assigning a value (maximum or minimum known value) to the outlier entry.
3) The percentage that outlier entries account for known entries is increased from 10% to 50% with an interval of 10%.
4) The outlier entities are only added into the training set. An example shown in Fig. 2 is illustrated for this method.

The results on all the datasets are presented in Fig. 3, where we see that although each model's MAE/RMSE increases as the percentage of outlier data increases, LFA-STSR still has lower MAE/RMSE than the other models on all the datasets. Besides, Fig. 3 shows that LFA-STSR is more robust than the other models to outlier data. For example, the MAE differences between LFA-STSR and the other six models (spatio-temporal low rank matrix approximation (ST-LRMA), batch reconstruction method of time-varying graph signals (BR-TVGS), LRDS, round-robin imputation (RRImpu), L_1-and-L_2-norm-oriented latent factor (L^3F), and TRSS) increase from 0.1861, 0.3702, 0.3528, 0.4371, 0.4323, and 0.36 to 1.1175, 0.9285, 0.8096, 1.2403, 0.8451, and 0.8150, respectively, when the percentage of outlier data increases from 0.1 to 0.5 on D3.

D. Ablation Studies (RQ. 3)

This set of experiments investigates how the L_1 and L_2 norms influence LFA-STSR model. To this end, we respec-

TABLE V
COMPARISON RESULTS OF RECOVERY ACCURACY ON D3 WITH DIFFERENT SAMPLING RATES

Sampling rate	Metric	ST-LRMA	BR-TVGS	LRDS	RRImpu	L^3F	TRSS	LFA-STSR
0.1	MAE	0.5270±0.0175	1.0013±0.0236	0.5480±0.0277	0.9137±0.0396	0.4481±0.0207	0.9269±0.0149	0.3661±0.0216
	RMSE	0.7323±0.0132	1.3882±0.0256	0.7408±0.0241	1.1445±0.0.406	1.1336±0.0218	1.2658±0.0296	0.4811±0.0242
0.2	MAE	0.3103±0.0115	0.6703±0.0296	0.3148±0.0114	0.6184±0.0163	0.3211±0.0141	0.6207±0.0180	0.2731±0.0166
	RMSE	0.4250±0.0167	0.9948±0.0311	0.4231±0.0225	0.8068±0.0296	0.4314±0.0161	0.8961±0.0168	0.3598±0.0243
0.3	MAE	0.3041±0.0187	0.4700±0.0167	0.2452±0.0149	0.4270±0.0199	0.2617±0.0158	0.4232±0.0242	0.2349±0.0159
	RMSE	0.4007±0.0195	0.7135±0.0177	0.3303±0.0192	0.5600±0.0165	0.3471±0.0222	0.6423±0.0168	0.3094±0.0172
0.4	MAE	0.2103±0.0290	0.3534±0.0214	0.2069±0.0229	0.4281±0.0243	0.2096±0.0262	0.3182±0.0103	0.1985±0.0168
	RMSE	0.2909±0.0128	0.5403±0.0180	0.2886±0.0264	0.5641±0.0102	0.2776±0.0157	0.4865±0.0188	0.2679±0.0139
0.5	MAE	0.1856±0.0168	0.2794±0.0188	0.1715±0.0274	0.4296±0.0170	0.1756±0.0221	0.2516±0.0274	0.1714±0.0109
	RMSE	0.2471±0.0103	0.4256±0.0216	0.2367±0.0280	0.5665±0.0165	0.2352±0.0202	0.3831±0.0168	0.2277±0.0108
0.6	MAE	0.1522±0.0196	0.2308±0.0197	0.1515±0.0164	0.4191±0.0173	0.1443±0.0253	0.2079±0.0130	0.1387±0.0160
	RMSE	0.2096±0.0178	0.3505±0.0242	0.2086±0.0168	0.5651±0.0196	0.1953±0.0153	0.3156±0.0196	0.189±0.0149
0.7	MAE	0.1368±0.0212	0.1940±0.0185	0.1359±0.0144	0.4185±0.0193	0.1343±0.0182	0.1746±0.0174	0.1289±0.0165
	RMSE	0.1884±0.0242	0.2823±0.0184	0.1868±0.0183	0.5652±0.0174	0.1825±0.0156	0.2542±0.0161	0.1764±0.0151
0.8	MAE	0.1268±0.0179	0.1734±0.0178	0.1260±0.0130	0.4064±0.0190	0.1255±0.0163	0.1563±0.0236	0.1207±0.0150
	RMSE	0.1753±0.0148	0.2438±0.0124	0.174±0.0149	0.5449±0.0112	0.1709±0.0126	0.2200±0.0277	0.1654±0.0155
0.9	MAE	0.1188±0.0197	0.1508±0.0141	0.1181±0.0128	0.4005±0.0161	0.1189±0.0187	0.1360±0.0396	0.1161±0.0153
	RMSE	0.1637±0.0230	0.2102±0.0177	0.1626±0.0181	0.5418±0.0194	0.1611±0.0147	0.1895±0.0296	0.1576±0.0145
Statistical Analysis	win/loss	18/0	18/0	18/0	18/0	18/0	18/0	Total 108/0
	p-value	**0.0001**	**0.0001**	**0.0001**	**0.0001**	**0.0001**	**0.0001**	—
	F-rank	3.61	6.33	2.72	6.39	2.67	5.28	**1.00**

TABLE VI
COMPARISON RESULTS OF RECOVERY ACCURACY ON D4 WITH DIFFERENT SAMPLING RATES

Sampling rate	Metric	ST-LRMA	BR-TVGS	LRDS	RRImpu	L^3F	TRSS	LFA-STSR
0.1	MAE	0.1687±0.0082	0.1501±0.0073	0.1501±0.0078	0.2151±0.0075	0.1565±0.0072	0.1520±0.0076	0.1463±0.0065
	RMSE	0.2639±0.0076	0.2236±0.0081	0.2230±0.0093	0.3087±0.0083	0.2356±0.0083	0.2382±0.0088	0.2195±0.0073
0.2	MAE	0.1531±0.0073	0.1311±0.0084	0.1309±0.0063	0.1414±0.0078	0.135±0.0074	0.1354±0.0063	0.1276±0.0069
	RMSE	0.2374±0.0067	0.1997±0.0091	0.1996±0.0079	0.2261±0.0096	0.2072±0.0092	0.2129±0.0052	0.1967±0.0086
0.3	MAE	0.1435±0.0042	0.1240 ±0.0059	0.1237±0.0076	0.1329±0.0056	0.1191±0.0069	0.1279±0.0074	0.1179±0.0091
	RMSE	0.2279±0.0053	0.1876±0.0078	0.1874±0.0079	0.2139±0.0073	0.1841±0.0073	0.2028±0.0086	0.1827±0.0085
0.4	MAE	0.1357±0.0068	0.1144±0.0086	0.1142±0.0075	0.1206±0.0066	0.1146±0.0068	0.1196±0.0099	0.1125±0.0068
	RMSE	0.2168±0.0085	0.1761±0.0084	0.1757±0.0086	0.1953±0.0076	0.1785±0.0067	0.1894±0.0064	0.1750±0.0072
0.5	MAE	0.1322±0.0061	0.1107±0.0091	0.1096±0.0062	0.1164±0.0069	0.1108±0.0092	0.1183±0.0065	0.1087±0.0061
	RMSE	0.2188±0.0081	0.1724±0.0066	0.1697±0.0082	0.1886±0.0086	0.1710±0.0084	0.1851±0.0093	0.1672±0.0059
0.6	MAE	0.1270±0.0053	0.1067±0.0075	0.1034±0.0075●	0.1082±0.0096	0.1042±0.0073	0.1142±0.0069	0.1037±0.0075
	RMSE	0.2163±0.0049	0.1707±0.0065	0.1639±0.0069	0.1696±0.0059	0.1612±0.0059	0.1825±0.0063	0.1606±0.0082
0.7	MAE	0.1197±0.0061	0.1027±0.0072	0.0999±0.0064●	0.1097±0.0067	0.1017±0.0062	0.1053±0.0065	0.1011±0.0069
	RMSE	0.1931±0.0054	0.1608±0.0031	0.1568±0.0093	0.1711±0.0076	0.1564±0.0073	0.1681±0.0071	0.1555±0.0076
0.8	MAE	0.1145±0.0075	0.0981±0.0065	0.0922±0.0074●	0.1064±0.0089	0.0984±0.0067	0.1011±0.0047	0.0969±0.0078
	RMSE	0.1916±0.0068	0.1585±0.0069	0.1488±0.0067	0.1667±0.0078	0.1488±0.0068	0.1627±0.0077	0.1471±0.0089
0.9	MAE	0.1057±0.0051	0.0925±0.0075	0.0878±0.0066●	0.1079±0.0067	0.0926±0.0045	0.0935±0.0059	0.0911±0.0069
	RMSE	0.1704±0.0047	0.1407±0.0074	0.1345±0.0049	0.1629±0.0079	0.1368±0.0046	0.1465±0.0062	0.1328±0.0075
Statistical Analysis	win/loss	18/0	18/0	14/4	18/0	18/0	18/0	104/4
	p-value	**0.0001**	**0.0001**	**0.0139**	**0.0001**	**0.0001**	**0.0001**	—
	F-rank	6.83	3.53	2.03	5.94	3.33	5.11	**1.22**

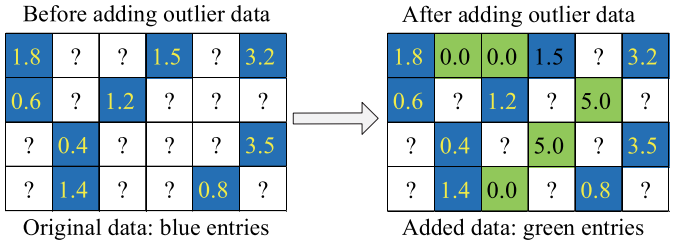

Fig. 2. Example of adding outlier data.

tively remove the L_1-norm-oriented or L_2-norm-oriented loss part of (7) to build two degraded variants of LFA-STSR, termed L_1-norm (only L_1-norm-oriented loss has remained) and L_2-norm (only L_2-norm-oriented loss has remained). The sampling rate is also set as 0.1. Fig. 4 shows the comparison results of the variants of LFA-STSR and itself. It can be found that the two degraded variants of LFA-STSR have their own advantages on the different datasets. None of them can guarantee the lower MAE/RMSE on all the datasets. However, by simultaneously combining both L_1 and L_2 norms to build the L_1-norm-and-L_2-norm-oritented loss as shown in (7), the advantages of both L_1 and L_2 norms are finely aggregated. Consequently, LFA-STSR achieves lower MAE/RMSE than the two degraded variants of LFA-STSR built on the L_1- or L_2-norm-oriented loss solely.

E. Effects by the Hyperparameters (RQ. 4)

In this section, we analyze the influences of k (latent feature dimension) and hyperparameters of z_1 and z_2 (controlling spatiotemporal correlation).

1) Latent Feature Dimension: According to prior studies [18], [19], [41], a larger latent feature dimension makes an LFA model have a better representation learning ability. Intuitively, more known data require a better LFA model to

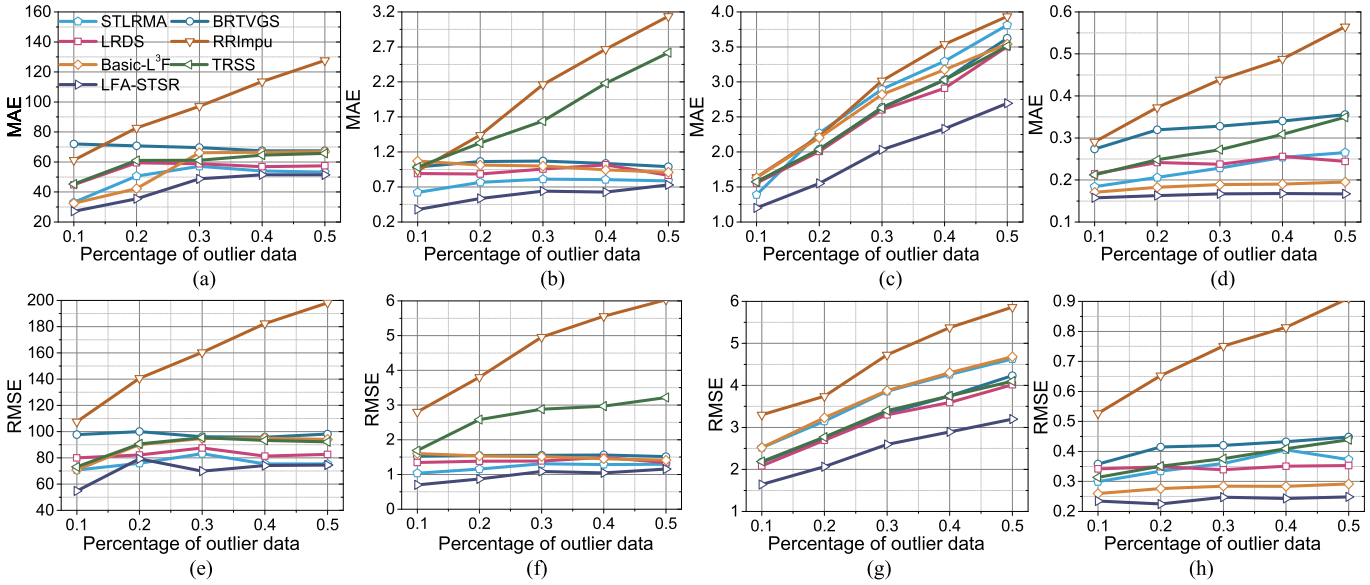

Fig. 3. Outlier data sensitivity tests on all the datasets. (a) MAE on D1. (b) MAE on D2. (c) MAE on D3. (d) MAE on D4. (e) RMSE on D1. (f) RMSE on D2. (g) RMSE on D3. (h) RMSE on D4.

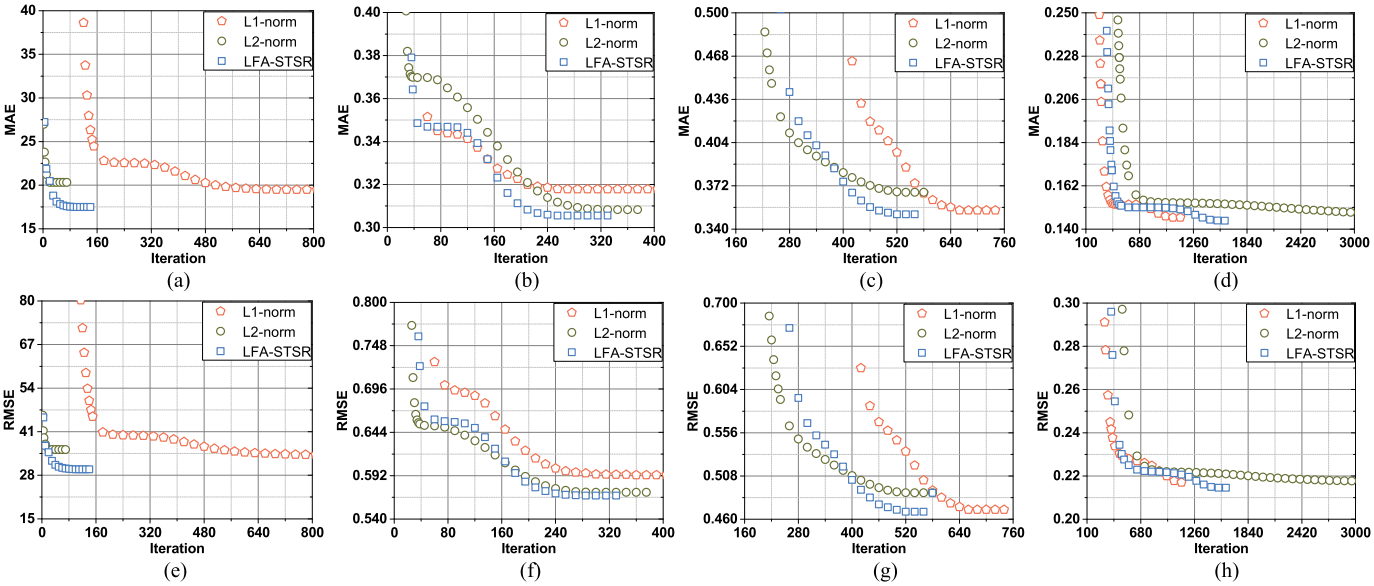

Fig. 4. Ablation testing results on all the datasets. (a) MAE on D1. (b) MAE on D2. (c) MAE on D3. (d) MAE on D4. (e) RMSE on D1. (f) RMSE on D2. (g) RMSE on D3. (h) RMSE on D4.

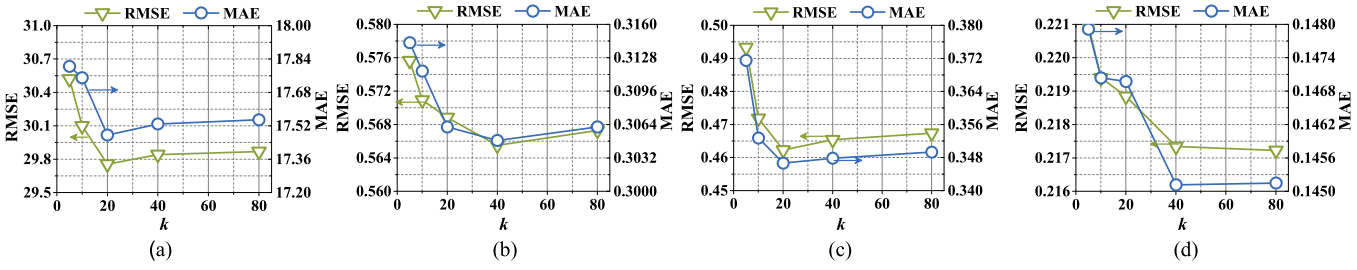

Fig. 5. MAE/RMSE of LFA-STSR as k increases from 5 to 80 on all the datasets. (a) D1, (b) D2, (c) D3, and (d) D4.

represent. Hence, this set of experiments set the sampling rate as 0.1. Fig. 5 records the results with different latent feature dimensions from 5 to 80. We see that MAE/RMSE keeps decreasing as k increases, which also substantiates that a larger k can improve LFA-STSR's recovery accuracy. However, the accuracy gain is not significant and even has worse accuracy when k increases over a certain threshold, e.g., 40. The reason is that overfitting occurs when a relatively

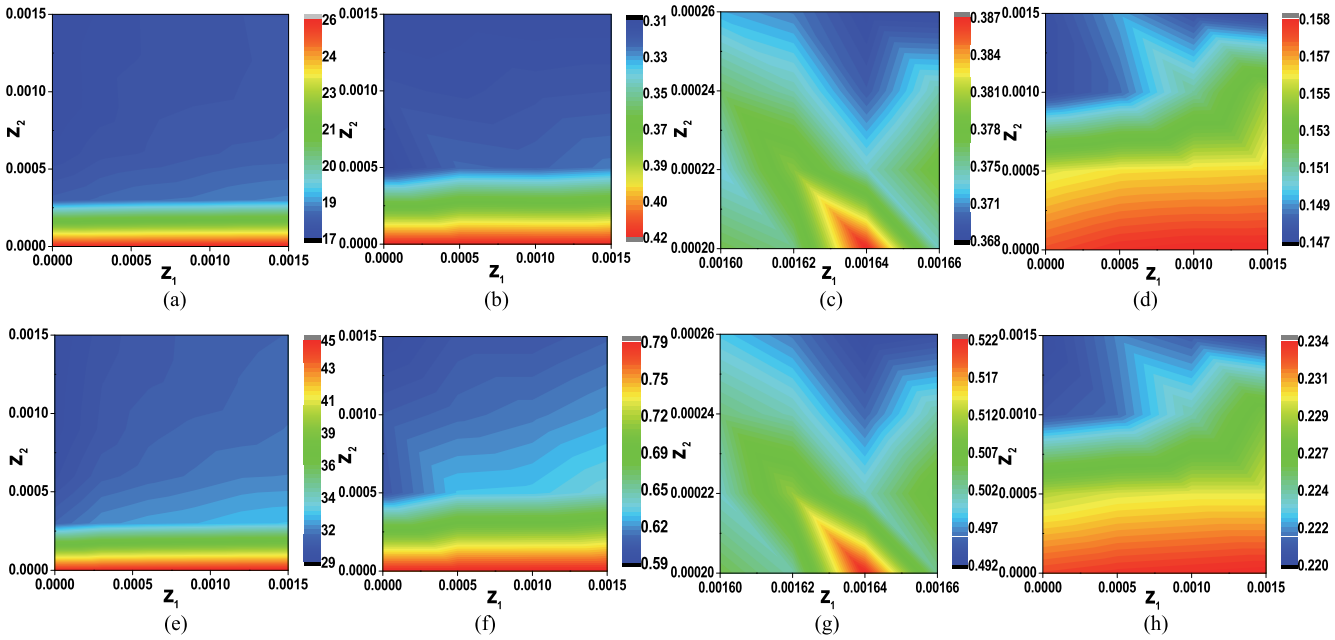

Fig. 6. MAE/RMSE of LFA-STSR with different z_1 and z_2 on all the datasets. (a) MAE on D1. (b) MAE on D2. (c) MAE on D3. (d) MAE on D4. (e) RMSE on D1. (f) RMSE on D2. (g) RMSE on D3. (h) RMSE on D4.

larger k is adopted to represent the relatively less-known data with a 0.1 sampling ratio. Besides, Section IV-E shows that a larger k increases the time complexity. Therefore, to balance the recovery accuracy and computational efficiency, we recommend setting $k = 40$ in practical applications.

2) Spatiotemporal Correlation: The sampling rate is set as 0.1. We perform a grid-based search to test the recovery accuracy of LFA-STSR with different z_1 and z_2. Fig. 6 records the results on all the datasets, where we see that both z_1 and z_2 influence LFA-STSR's recovery accuracy. As z_1 and z_2 increase, MAE/RMSE decreases till they reach their optimal thresholds and then increase again. For example, the optimal values of z_1 and z_2 are 0.0016 and 0.0002, respectively. These results substantiate that the spatiotemporal correlation is important for LFA-STSR to improve recovery accuracy. Besides, we note that z_1 has a small influence on D1 and D2. The reason may be that the spatial correlation of the two datasets is relatively weak on D1 and D2.

VI. Conclusion

This article proposes a LFA based STSR model, named LFA-STSR, for data recovery in WSNs. Its main idea is twofold: 1) the spatiotemporal correlation is incorporated into an LFA model as the regularization constraint to improve its recovery accuracy and 2) the L_1-norm is aggregated into the loss part of an LFA model to improve its robustness to outliers. With such a design, our LFA-STSR model possesses the merits of both high accuracy and robustness during data recovery in WSNs. In the experiments, LFA-STSR is evaluated on four real-world WSNs datasets. The results demonstrate that LFA-STSR significantly outperforms the related state-of-the-art models in terms of both recovery accuracy and robustness to outliers.

According to Fig. 6, it is evident that the hyperparameters z_1 and z_2, which control the spatiotemporal correlation, have a significant impact on the recovery accuracy of LFA-STSR. Furthermore, it is important to note that these hyperparameters are data-dependent. Presently, they are tuned by grid-search on the validation data partially sampled from training data. However, such grid-search is time-consuming on large-scale datasets, which reduces practicality in big data-related applications. Therefore, we plan to make z_1 and z_2 self-adaptive based on Taguchi's methods, advanced grid-search, and intelligent optimization methods [67], [68], [69] in the future.

References

[1] Z. Lv, L. Qiao, A. Kumar Singh, and Q. Wang, "AI-empowered IoT security for smart cities," *ACM Trans. Internet Technol.*, vol. 21, no. 4, pp. 1–21, Nov. 2021.

[2] Z. Zhao, D. F. Wang, Y. Hou, T. Itoh, and R. Maeda, "Noninvasive passive measurement of multiphase currents in IoT," *IEEE Trans. Ind. Electron.*, vol. 68, no. 12, pp. 12860–12870, Dec. 2021.

[3] Y. Peng, W. Qiao, and L. Qu, "Compressive sensing-based missing-data-tolerant fault detection for remote condition monitoring of wind turbines," *IEEE Trans. Ind. Electron.*, vol. 69, no. 2, pp. 1937–1947, Feb. 2022.

[4] B. Li, Y. Zou, J. Zhu, and W. Cao, "Impact of hardware impairment and co-channel interference on security-reliability trade-off for wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7011–7025, Nov. 2021.

[5] S. A. Putra, B. R. Trilaksono, M. Riyansyah, and D. S. Laila, "Multiagent architecture for bridge capacity measurement system using wireless sensor network and weight in motion," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–14, 2021.

[6] Y. Gong et al., "Self-powered wireless sensor node for smart railway axle box bearing via a variable reluctance energy harvesting system," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.

[7] A. Grover and B. Lall, "A recursive method for estimating missing data in spatio-temporal applications," *IEEE Trans. Ind. Informat.*, vol. 18, no. 4, pp. 2714–2723, Apr. 2022.

[8] J. Song, S. Xia, J. Wang, M. Patel, and D. Chen, "Uncertainty quantification of hyperspectral image denoising frameworks based on sliding-window low-rank matrix approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 3065570.

[9] S. Sarkar and R. R. Sahay, "A non-local superpatch-based algorithm exploiting low rank prior for restoration of hyperspectral images," *IEEE Trans. Image Process.*, vol. 30, pp. 6335–6348, 2021.

[10] T. Xie, S. Li, and J. Lai, "Adaptive rank and structured sparsity corrections for hyperspectral image restoration," *IEEE Trans. Cybern.*, vol. 52, no. 9, pp. 8729–8740, Sep. 2022.

[11] X. Luo, Y. Yuan, S. Chen, N. Zeng, and Z. Wang, "Position-transitional particle swarm optimization-incorporated latent factor analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 8, pp. 3958–3970, Aug. 2022, doi: 10.1109/TKDE.2020.3033324.

[12] X. He, J. Tang, X. Du, R. Hong, T. Ren, and T.-S. Chua, "Fast matrix factorization with nonuniform weights on missing data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 8, pp. 2791–2804, Aug. 2020.

[13] J. Wang, H. Han, H. Li, S. He, P. Kumar Sharma, and L. Chen, "Multiple strategies differential privacy on sparse tensor factorization for network traffic analysis in 5G," *IEEE Trans. Ind. Informat.*, vol. 18, no. 3, pp. 1939–1948, Mar. 2022.

[14] H. Li, K. Li, J. An, and K. Li, "An online and scalable model for generalized sparse nonnegative matrix factorization in industrial applications on multi-GPU," *IEEE Trans. Ind. Informat.*, vol. 18, no. 1, pp. 437–447, Jan. 2022.

[15] K. Xie et al., "Accurate recovery of internet traffic data: A sequential tensor completion approach," *IEEE/ACM Trans. Netw.*, vol. 26, no. 2, pp. 793–806, Apr. 2018.

[16] J.-Q. Lin, H.-C. Wu, and S.-C. Chan, "A new regularized recursive dynamic factor analysis with variable forgetting factor and subspace dimension for wireless sensor networks with missing data," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021.

[17] C. Leng, H. Zhang, G. Cai, I. Cheng, and A. Basu, "Graph regularized lp smooth non-negative matrix factorization for data representation," *IEEE/CAA J. Autom. Sinica*, vol. 6, no. 2, pp. 584–595, Mar. 2019.

[18] X. Luo, H. Wu, Z. Wang, J. Wang, and D. Meng, "A novel approach to large-scale dynamically weighted directed network representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 9756–9773, Dec. 2022.

[19] J. Pan and N. Gillis, "Generalized separable nonnegative matrix factorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 5, pp. 1546–1561, May 2021.

[20] W. Shi et al., "Effective prediction of missing data on apache spark over multivariable time series," *IEEE Trans. Big Data*, vol. 4, no. 4, pp. 473–486, Dec. 2018.

[21] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.

[22] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.

[23] A. Ortega, P. Frossard, J. Kovacevic, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, vol. 106, no. 5, pp. 808–828, May 2018.

[24] X. Zhu and M. Rabbat, "Graph spectral compressed sensing for sensor networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 2865–2868.

[25] X. Shi, H. Feng, M. Zhai, T. Yang, and B. Hu, "Infinite impulse response graph filters in wireless sensor networks," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1113–1117, Aug. 2015.

[26] X. Zhu et al., "Similarity-maintaining privacy preservation and location-aware low-rank matrix factorization for QoS prediction based web service recommendation," *IEEE Trans. Services Comput.*, vol. 14, no. 3, pp. 889–902, May 2021.

[27] D. Wu, X. Luo, M. Shang, Y. He, G. Wang, and X. Wu, "A data-characteristic-aware latent factor model for web service QoS prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 6, pp. 2525–2538, Aug. 2022.

[28] X. Mao and Y. Gu, "Time-varying graph signals reconstruction," in *Vertex-Frequency Analysis of Graph Signals*. Cham, Switzerland: Springer, 2019, pp. 293–316.

[29] Y. Liu, W. Guo, K. You, L. Zhao, T. Peng, and W. Wang, "Graph learning for spatiotemporal signals with long- and short-term characterization," *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 6, pp. 699–713, 2020.

[30] H. Wang, H. Qiao, J. Lin, R. Wu, Y. Liu, and Q. Dai, "Model study of transient imaging with multi-frequency time-of-flight sensors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3523–3539, Oct. 2021.

[31] T. Wang and D. J. Cook, "SMRT: Multi-resident tracking in smart homes with sensor vectorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 8, pp. 2809–2821, Aug. 2021.

[32] J. Miao and K. I. Kou, "Color image recovery using low-rank quaternion matrix completion algorithm," *IEEE Trans. Image Process.*, vol. 31, pp. 190–201, 2022.

[33] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*. Cambridge, U.K.: Cambridge Univ. Press, 2006.

[34] *Sea Surface Temperature (SST) V2, Physical Sciences Division, EarthSystem Research Laboratory, National Oceanic and Atmospheric Administration*. Accessed: Dec. 18, 2015. [Online]. Available: http://www.esrl.noaa.gov/psd/data/grided/data.noaa.oisst.v2.html

[35] (May 19, 2016). *Air Quality Data, United States Environmental Protection Agency*. [Online]. Available: https://www.epa.gov/outdoor-air-quality-data

[36] P. Xu, W. Ruan, Q. Z. Sheng, T. Gu, and L. Yao, "Interpolating the missing values for multi-dimensional spatial–temporal sensor data: A tensor SVD approach," in *Proc. 14th EAI Int. Conf. Mobile Ubiquitous Syst., Comput., Netw. Services*, 2018, pp. 442–451.

[37] (2016). *Beijing $PM_{2.5}$ Data, Beijing Municipal Ecological and Environmental Monitoring Center*. [Online]. Available: https://quotsoft.net/air/

[38] X. Piao, Y. Hu, Y. Sun, B. Yin, and J. Gao, "Correlated spatio-temporal data collection in wireless sensor networks based on low rank matrix approximation and optimized node sampling," *Sensors*, vol. 14, no. 12, pp. 23137–23158, Dec. 2014.

[39] K. Qiu, X. Mao, X. Shen, X. Wang, T. Li, and Y. Gu, "Time-varying graph signal reconstruction," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 6, pp. 870–883, Sep. 2017.

[40] X. Mao, K. Qiu, T. Li, and Y. Gu, "Spatio-temporal signal recovery based on low rank and differential smoothness," *IEEE Trans. Signal Process.*, vol. 66, no. 23, pp. 6281–6296, Dec. 2018.

[41] D. Wu, M. Shang, X. Luo, and Z. Wang, "An L_1-and-L_2-norm-oriented latent factor model for recommender systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 10, pp. 5775–5788, Oct. 2022, doi: 10.1109/TNNLS.2021.3071392.

[42] B. Muzellec, J. Josse, C. Boyer, and M. Cuturi, "Missing data imputation using optimal transport," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 7130–7140.

[43] Y. Ma, Z. Wang, H. Yang, and L. Yang, "Artificial intelligence applications in the development of autonomous vehicles: A survey," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 2, pp. 315–329, Mar. 2020.

[44] H. Yan, J. Wang, H. Zhang, H. Shen, and X. Zhan, "Event-based security control for stochastic networked systems subject to attacks," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 50, no. 11, pp. 4643–4654, Nov. 2020.

[45] X. Zhang, J. He, Y. Li, Y. Chi, and Y. Zhou, "Recovery of corrupted data in wireless sensor networks using tensor robust principal component analysis," *IEEE Commun. Lett.*, vol. 25, no. 10, pp. 3389–3393, Oct. 2021.

[46] L. Yang, Y. C. Eldar, H. Wang, K. Kang, and H. Qian, "An ADMM-Net for data recovery in wireless sensor networks," in *Proc. 28th Eur. Signal Process. Conf. (EUSIPCO)*, Jan. 2021, pp. 1712–1716.

[47] Y. Cherapanamjeri, K. Gupta, and P. Jain, "Nearly optimal robust matrixcompletion," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 797–805.

[48] Y. Xie, D. Tao, W. Zhang, Y. Liu, L. Zhang, and Y. Qu, "On unifying multi-view self-representations for clustering by tensor multi-rank minimization," *Int. J. Comput. Vis.*, vol. 126, no. 11, pp. 1157–1179, Nov. 2018.

[49] H. Mansour, D. Tian, and A. Vetro, "Factorized robust matrix completion," in *Handbook of Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing*, vol. 5. Boca Raton, FL, US: CRC Press, 2016.

[50] Q. Shi, H. Lu, and Y.-M. Cheung, "Rank-one matrix completion with automatic rank estimation via L_1-norm regularization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4744–4757, Oct. 2018.

[51] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Amer. Stat. Assoc.*, vol. 58, no. 301, pp. 13–30, Mar. 1963.

[52] J. Fan, L. Ding, Y. Chen, and M. Udell, "Factor group-sparse regularization for efficient low-rank matrix recovery," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 5105–5115.

[53] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and Bregman iterative methods for matrix rank minimization," *Math. Program.*, vol. 128, nos. 1–2, pp. 321–353, Jun. 2011.
[54] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, Jan. 2006.
[55] F. Shang, J. Cheng, Y. Liu, Z.-Q. Luo, and Z. Lin, "Bilinear factor matrix norm minimization for robust PCA: Algorithms and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 9, pp. 2066–2080, Sep. 2018.
[56] C. Xu, Z. Lin, and H. Zha, "A unified convex surrogate for the schatten-p norm," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 926–932.
[57] H. Kong, X. Xie, and Z. Lin, "T-schatten-p norm for low-rank tensor recovery," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 6, pp. 1405–1419, Dec. 2018.
[58] X. Han, J. Wu, L. Wang, Y. Chen, L. Senhadji, and H. Shu, "Linear total variation approximate regularized nuclear norm optimization for matrix completion," *Abstract Appl. Anal.*, vol. 2014, pp. 1–8, May 2014.
[59] T. Yokota, Q. Zhao, and A. Cichocki, "Smooth PARAFAC decomposition for tensor completion," *IEEE Trans. Signal Process.*, vol. 64, no. 20, pp. 5423–5436, Oct. 2016.
[60] J. H. Giraldo, A. Mahmood, B. Garcia-Garcia, D. Thanou, and T. Bouwmans, "Reconstruction of time-varying graph signals via sobolev smoothness," *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 8, pp. 201–214, 2022.
[61] X. Chang and Y. Yang, "Semisupervised feature analysis by mining correlations among multiple tasks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2294–2305, Oct. 2017.
[62] M. Luo, X. Chang, L. Nie, Y. Yang, A. G. Hauptmann, and Q. Zheng, "An adaptive semisupervised feature analysis for video semantic recognition," *IEEE Trans. Cybern.*, vol. 48, no. 2, pp. 648–660, Feb. 2018.
[63] M. Luo, F. Nie, X. Chang, Y. Yang, A. G. Hauptmann, and Q. Zheng, "Adaptive unsupervised feature selection with structure regularization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 944–956, Apr. 2018.
[64] F. Nie, H. Huang, X. Cai, and C. Ding, "Efficient and robust feature selection via joint $\ell_{2,1}$-norms minimization," in *Proc. 24th Annu. Conf. Neural Inf. Process. Syst.*, 2010, pp. 1813–1821.
[65] Z. Wei et al., "STGSA: A novel spatial–temporal graph synchronous aggregation model for traffic prediction," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 1, pp. 226–238, Jan. 2023.
[66] H. Huang, G. Zhou, N. Liang, Q. Zhao, and S. Xie, "Diverse deep matrix factorization with hypergraph regularization for multiview data representation," *IEEE/CAA J. Autom. Sinica*, early access, Oct. 4, 2022, doi: 10.1109/JAS.2022.105980.
[67] S. Gao, M. Zhou, Y. Wang, J. Cheng, H. Yachi, and J. Wang, "Dendritic neuron model with effective learning algorithms for classification, approximation, and prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 2, pp. 601–614, Feb. 2019.
[68] P. Zhang, S. Shu, and M. Zhou, "An online fault detection model and strategies based on SVM-grid in clouds," *IEEE/CAA J. Autom. Sinica*, vol. 5, no. 2, pp. 445–456, Mar. 2018.
[69] C. Chen, N. Lu, B. Jiang, and C. Wang, "A risk-averse remaining useful life estimation for predictive maintenance," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 2, pp. 412–422, Feb. 2021.
[70] Y. Xie et al., "Robust kernelized multiview self-representation for subspace clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 868–881, Feb. 2021.

Di Wu (Member, IEEE) received the Ph.D. degree from the Chongqing Institute of Green and Intelligent Technology (CIGIT), Chinese Academy of Sciences (CAS), Chongqing, China, in 2019.

He joined CIGIT. He is currently a Professor at the College of Computer and Information Science, Southwest University, Chongqing. He has over 80 publications, including more than 20 IEEE/ACM TRANSACTIONS articles and several conference papers on AAAI, ICDM, WWW, IJCAI, ECML-PKDD, etc. His research interests include machine learning and data mining.

Dr. Wu is serving as an Associate Editor for *Neurocomputing* and *Frontiers in Neurorobotics*.

Zechao Li (Senior Member, IEEE) received the B.E. degree from the University of Science and Technology of China, Hefei, China, in 2008, and the Ph.D. degree from the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2013.

He is currently a Professor at the Nanjing University of Science and Technology, Nanjing, China. His research interests include big media analysis, computer vision, etc.

Dr. Li was a recipient of the Best Paper Award in ACM Multimedia in Asia 2020 and the Best Student Paper Award in ICIMCS 2018. He is currently serving as an Associate Editor for IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS.

Zhikai Yu (Student Member, IEEE) received the B.E. degree in electronic information science and technology from Minnan Normal University, Zhangzhou, China, in 2019. He is currently pursuing the M.S. degree in computer technology with the Chongqing University of Posts and Telecommunications, Chongqing, China.

He has been a Visiting Student at CIGIT, CAS, Chongqing, since July 2021. His current research interests include machine learning and data mining.

Yi He (Member, IEEE) received the B.E. degree from the Harbin Institute of Technology, Harbin, China, in 2013, and the Ph.D. degree in computer science from the University of Louisiana at Lafayette, Lafayette, LA, USA, in 2020.

He is an Assistant Professor of computer science at ODU, Norfolk, VA, USA. His research focus lies broadly in data mining and machine learning and specifically in online learning, data stream analytics, and graph theory. His research outcomes have appeared in top venues, e.g., AAAI, IJCAI, WWW, ICDM, SDM, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, to name a few.

Dr. He has served as multiple roles in the Data Mining and Machine Learning Community, including the Registration Chair of ICDM 2022, the Session Chair of ICMR 2022, and the (T)PC member of AAAI'20,'21,'22, IJCAI'22, ICDM'22, CIKM'22, ECMLPKDD'22, and referees for prestigious journals such as IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON MOBILE COMPUTING, and IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS.

Xin Luo (Senior Member, IEEE) received the B.S. degree in computer science from the University of Electronic Science and Technology of China, Chengdu, China, in 2005, and the Ph.D. degree in computer science from Beihang University, Beijing, China, in 2011.

He is currently a Professor of data science and computational intelligence with the College of Computer and Information Science, Southwest University, Chongqing, China. He has authored or coauthored over 240 articles (including over 120 IEEE TRANSACTIONS articles) in the areas of his interests. His research interests focus on big data analysis and graph learning.

Dr. Luo was a recipient of the Outstanding Associate Editor Award from IEEE/CAA JOURNAL OF AUTOMATICA SINICA in 2020. He is currently serving as a Deputy-Editor-in-Chief for IEEE/CAA JOURNAL OF AUTOMATICA SINICA and an Associate Editor for IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS.